Τμήμα Μηχανικών Η/Υ , Τηλεπικοινωνιών και Δικτύων

Πανεπιστήμιο Θεσσαλίας

"Αξιολόγηση Ποιότητας Μηνυμάτων Ηλεκτρονικού Ταχυδρομείου με Χρήση Αλγορίθμων HASH"

Φοιτητής: Βαραλής Αργύρης

Επιβλέπων: καθ. Λέανδρος Τασιούλας

Φεβρουάριος 2011

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ. Τασιούλα Λέανδρο Καθηγητή Τηλεπικοινωνιών και Δικτύων στο Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του Πανεπιστημίου Θεσσαλίας για την ανάθεση του θέματος.

Επίσης θα ήθελα να ευχαριστήσω τον κ. Σπύρο Κοψιδά υπ. Διδάκτωρ του Πανεπιστημίου Θεσσαλίας για τις υποδείξεις, την καθοδήγηση και την υπομονή που έδειξε όλο αυτό τον χρόνο.

Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου για την οικονομική και κυρίως ηθική υποστήριξη όλα αυτά τα χρόνια.

Περιεχόμενα

Εισαγωγή (Περίληψη) 3
1. Ηλεκτρονικό Ταχυδρομείο και Ανεπιθύμητη Αλληλογραφία (SPAM) 4
1.1Τεχνικές spam 6
1.2.Ποιοι είναι οι Spammers 7
2.Υπάρχουσες Λύσεις Αντιμετώπισης SPAM (State-of-the-Art) 9
2.1 Γενικές πληροφορίες 9
2.2.Προστασία από ανεπιθύμητη αλληλογραφία(spam) 10
2.2.1 Για τους απλούς χρήστες 10
2.2.2. Μέθοδοι φιλτραρίσματος 12
2.2.3Περιορισμοί φίλτρων 14
2.2.4.Μειονεκτήματα 15
2.2.5.Για τους διαχειριστές εξυπηρετητών ηλεκτρονικού ταχυδρομείου (mail servers) 16
3.Αλγόριθμοι HASH 18
3.1.Αλγόριθμοι hash 18
3.2.Αλγόριθμος MD5 hash 18
4.Ανάλυση Συστήματος Anti-SPAM 19
4.1.Γενικές πληροφορίες συστήματος 19
4.2.Αναλυτική περιγραφή εφαρμογής 20
5.Μελλοντικές Επεκτάσεις 25
6. Συμπεράσματα 25
7.Βιβλιογραφία 26

Εισαγωγή (Περίληψη)

Το φαινόμενο του e-mail spam είναι ένα φαινόμενο που υπάρχει εδώ και πολλά χρόνια και παρόλο που είναι ανεπιθύμητο και γίνονται μεγάλες προσπάθειες για την καταπολέμηση του, κάθε χρόνο το ποσοστό αποστολής spam e-mails αυξάνεται αντί να μειώνεται.Υπάρχουν διάφοροι τρόποι αντιμετώπισής του, όμως οι spammers βρίσκουν τρόπους για την αποστολή του. Σε αυτή την εργασία αναλύεται μια νέα μέθοδος αντιμετώπισης του spam, που βασίζεται στους απλούς χρήστες αλλά και στους διαχειριστές των mail server για την αποτελεσματική λειτουργία του. Είναι πολύ απλή και γρήγορη μέθοδος αντιμετώπισης για τον χρήστη. Σκοπός του είναι να περιορίσει το spam και να ληφθεί από όσο λιγότερους χρήστες γίνεται.

1. Ηλεκτρονικό Ταχυδρομείο και Ανεπιθύμητη Αλληλογραφία (SPAM)

Η υπηρεσία ηλεκτρονικού ταχυδρομείου επιτρέπει σε κάθε χρήστη να λαμβάνει και να στέλνει ηλεκτρονικά μηνύματα, ελεγμένα για την ύπαρξη γνωστών ιών, μέσω της προσωπικής του διεύθυνσης, της μορφής: πχ username@domain.auth.gr. Οι χρήστες μπορούν να δίνουν τη διεύθυνση αυτή σε οποιονδήποτε επιθυμούν να επικοινωνήσει μαζί τους. Για κάθε χρήστη, ο κεντρικός διακομιστής ηλεκτρονικού ταχυδρομείου (mail server) διατηρεί μία θυρίδα όπου αποθηκεύονται τα ηλεκτρονικά μηνύματα του χρήστη, μέχρι εκείνος να τα διαβάσει και να τα διαγράψει.

Spam είναι η μαζική αποστολή μεγάλου αριθμού μηνυμάτων που απευθύνονται σε ένα σύνολο παραληπτών του Διαδικτύου χωρίς αυτοί να το επιθυμούν και χωρίς να έχουν συνειδητά προκαλέσει την αλληλογραφία με τον εν λόγω αποστολέα. Το spam συχνά έχει τη μορφή ενημερωτικών ή διαφημιστικών μηνυμάτων για προϊόντα ή υπηρεσίες τα οποία φθάνουν στο γραμματοκιβώτιό μας χωρίς να έχουμε ζητήσει αυτήν την πληροφόρηση. Αυτή η αλληλογραφία λοιπόν, μπορεί να χαρακτηριστεί ως ανεπιθύμητη.

Τα κυριότερα χαρακτηριστικά του spam μπορούν να συνοψιστούν στα ακόλουθα σημεία:

- **Απρόκλητο:** Η επικοινωνία που επιχειρείται είναι απρόκλητη, δηλαδή δεν υπάρχει κάποια σχέση μεταξύ παραλήπτη και αποστολέα που θα δικαιολογούσε ή θα προκαλούσε την επικοινωνία αυτή.

- **Εμπορικό:** Πολλές φορές το spam αφορά στην αποστολή μηνυμάτων εμπορικού σκοπού για την προβολή και τη διαφήμιση προϊόντων και υπηρεσιών, με σκοπό την προσέλκυση πελατών και την πραγματοποίηση πωλήσεων.
- **Μαζικό:** Το spam συνίσταται στη μαζική αποστολή μεγάλων ποσοτήτων μηνυμάτων από τον αποστολέα σε ένα μεγάλο πλήθος παραληπτών.

Ο κύριος λόγος που στέλνονται τα spam-mails είναι επειδή είναι ο πιο φτηνός τρόπος διαφήμισης. Το κόστος του δε συγκρίνεται με το κόστος του άμεσου marketing. Δεν υπάρχουν έξοδα εκτύπωσης, ούτε τηλεπικοινωνιακά κόστη, σχεδόν μηδαμινό κόστος δημιουργίας του, και κανένας δεν ελέγχει το περιεχόμενο τους. ISPs που επιτρέπουν στους χρήστες του να στέλνουν spam μπορεί να μπούνε σε μαύρες λίστες και τα email των πελατών του να μπλοκάρονται από άλλους ISPs.

Το email spam αυξάνεται σταθερά από τις αρχές του '90. Τα Bot nets(κακόβουλο λογισμικό) και τα δίκτυα μολυσμένων υπολογιστών από ιούς, ευθύνονται για την αποστολή περίπου 80% των spam mails. Οι spammers συλλέγουν διευθύνσεις ηλεκτρονικού ταχυδρομείου από τα chat rooms, από ιστοσελίδες, από καταλόγους πελατών, από newsgroups και από ιούς που «ψαρεύουν» e-mails και τα πουλούν στους spammers. Τα spam υπολογίζονται κατά μέσο όρο στο 78% του ηλεκτρονικού ταχυδρομείου που στέλνεται. Σύμφωνα με το Message Anti-Abuse Working Group το ποσό του ηλεκτρονικού ταχυδρομείου spam ήταν μεταξύ 88-92% των μηνυμάτων ηλεκτρονικού ταχυδρομείου που στάλησαν το πρώτοο εξάμηνο του 2010.

1.1Τεχνικές spam

Υπάρχουν διάφορες τεχνικές για την αποστολή spam mails:

Προσάρτηση, για παράδειγμα κάποιος έμπορος έχει μια βάση δεδομένων με ονόματα, διευθύνσεις, τηλέφωνα, μπορεί να πληρώσει για να αντιστοιχίσει τη βάση του σε μια εξωτερική βάση που περιέχει άλλες διευθύνσεις ηλεκτρονικού ταχυδρομείου. Με αυτό το τρόπο ο έμπορος έχει τη δυνατότητα να στείλει e-mails σε πρόσωπα που δεν έχουν ζητήσει τέτοιου είδους e-mails.

Spam με εικόνα, είναι μια μέθοδος κατά την οποία το κείμενο του μηνύματος αποθηκεύεται ως εικόνα GIF ή JPEG. Αυτό αποτρέπει τα διάφορα φίλτρα, που βασίζονται στο κείμενο, να μπλοκάρουν το μήνυμα.

Το κενό spam στερείται διαφήμισης. Συχνά το σώμα του μηνύματος λείπει, καθώς και το θέμα του μηνύματος. Το κενό spam μπορεί να προέρχεται από άλλου είδους επιθέσεις, όπως την απόκτηση έγκυρων e-mail. Μερικά κενά spam μπορεί να φαίνονται κενά αλλά στη πραγματικότητα να μην είναι. Χρησιμοποιούν HTML κώδικα για να κατεβάσουν άλλα αρχεία.

Οπισθοδιασπορά είναι μια παρενέργεια του email spam, των ιών και των worms όπου οι mail servers που λαμβάνουν spam και άλλα mails, στέλνουν μηνύματα αναπήδησης σε αθώους χρήστες. Αυτό συμβαίνει γιατί ο αποστολέας του αρχικού μηνύματος είναι φτιαγμένος να περιέχει τη διεύθυνση του θύματος.

1.2.Ποιοι είναι οι Spammers

Το 90% του spam που παράγεται σε παγκόσμιο επίπεδο, προέρχεται από 200 περίπου ομάδες spammers (κυρίως από τις ΗΠΑ, την Κίνα και την Νότιο Κορέα). Οι ομάδες αυτές διαθέτουν άριστη τεχνογνωσία και χρησιμοποιούν προηγμένο λογισμικό για την υλοποίηση των στόχων τους. Το υπόλοιπο 10% (που και πάλι μεταφράζεται σε δισεκατομμύρια spam-mails) δημιουργείται από μικρές επιχειρήσεις που δεν εφαρμόζουν τους κανόνες του e-mail marketing, εκμεταλλευόμενοι την αφέλεια ή την άγνοια των χρηστών, από κάθε λογής και διαμετρήματος απατεώνες. Αυτό το 10% των συνολικών ποσοτήτων spam έχει υπολογιστεί ότι προέρχεται από 100.000 περίπου spammers.

Οι 200 οργανωμένες ομάδες spammer και ένα ποσοστό από τους υπόλοιπους, για να αποφύγουν τις νομικές και οικονομικές συνέπειες των πράξεών τους, πλαστογραφούν τις e-mail διευθύνσεις τους ώστε να μην είναι δυνατός (ή να καθίσταται εξαιρετικά δυσχερής) ο

εντοπισμός τους. Τις περισσότερες φορές, το e-mail και το όνομα που παρουσιάζεται ως ο αποστολέας των ποσοτήτων spam-mail είναι κάποιος, εντελώς αθώος, χρήστης του διαδικτύου, του οποίου χρησιμοποιήθηκε εν αγνοία του η e-mail διεύθυνσή του. Αυτό έχει σαν συνέπεια να μπλοκάρει ο λογαριασμός του από τις επιστροφές των ανεπίδοτων spam-mail (κάθε αποστολή spam , περιλαμβάνει χιλιάδες ή και εκατομμύρια διευθύνσεις, πολλές από τις οποίες έχουν καταργηθεί) ή/και από τις εκατοντάδες/χιλιάδες διαμαρτυρίες των παραληπτών του spam. Με λίγα λόγια, μπορεί κάποιος spammer να χρησιμοποιήσει το e-mail ενός χρήστη με πλήρη άγνοια του και να εμφανίζετε ως ο αποστολέας μιας τεράστιας ποσότητας spam. Αυτό αποτελεί την πιο συνηθισμένη τακτική των spammers.

Οι spammers συλλέγουν διευθύνσεις ηλεκτρονικού ταχυδρομείου από τα chat rooms, από ιστοσελίδες, από καταλόγους πελατών, από newsgroups και από ιούς που «ψαρεύουν» e-mails και τα πουλούν στους spammers.

Τα πιο συνηθισμένα προϊόντα που διαφημίζονται, σύμφωνα με τις πληροφορίες που συντάσσονται από Commtouch Software Ltd., με το ηλεκτρονικό ταχυδρομείο spam για το πρώτο τρίμηνο του 2010 μπορεί να αναλυθεί ως εξής:

E-Mail Spam by Topic

Pharmacy	81%
Replica	5.40%
Enhancers	2.30%
Phishing	2.30%
Degrees[16]	1.30%
Casino	1%
Weight Loss	0.40%
Other	6.30%

2.Υπάρχουσες Λύσεις Αντιμετώπισης SPAM (State-of-the-Art)

2.1 Γενικές πληροφορίες

Όλοι οι απλοί χρήστες του ηλεκτρονικού ταχυδρομείου έχουν σίγουρα ενοχληθεί από την αποστολή τέτοιων απρόκλητών και ενοχλητικών μηνυμάτων. Ιδιαίτερο πρόβλημα αντιμετωπίζουν οι χρήστες που χρησιμοποιούν σε μεγάλα διαστήματα της μέρας το ηλεκτρονικό ταχυδρομείο και είναι αναγκασμένοι να σβήνουν όλη αυτή την ανεπιθύμητη αλληλογραφία. Τα μηνύματα αυτά για αρκετούς χρήστες φθάνουν να είναι πολλές φορές εκατοντάδες σε μια μέρα. Η αναγκαιότητα για την αντιμετώπιση του Spam εντοπίζεται στα ακόλουθα σημεία:

- **Είναι φαινόμενο δυσάρεστο**, ενοχλητικό και απαράδεκτο από τους παραλήπτες.
- Πολλές φορές προβάλει αμφίβολης ποιότητας προϊόντα και υπηρεσίες, ενώ συνηθισμένη είναι και η προβολή ύποπτων οικονομικών δραστηριοτήτων τύπου πυραμίδων κλπ. 'Άλλα μηνύματα περιέχουν ή διαφημίζουν σεξουαλικό περιεχόμενο.
- **Οδηγεί σε κατάχρηση πόρων του Διαδικτύου.** Η κατάχρηση αυτή επιβαρύνει τα δίκτυα με κατανάλωση εύρους ζώνης, αποθηκευτικών και υπολογιστικών πόρων στα κεντρικά συστήματα διανομής αλληλογραφίας (e-mail servers). Αντίστοιχα

προβλήματα προκαλεί στην πρόσβαση και στα συστήματα των χρηστών.

- **Θέτει σε κίνδυνο την ασφάλεια και την αξιοπιστία του διαδικτύου.** Οι spammers βρίσκονται σε συνεχή αναζήτηση συστημάτων τα οποία θα μπορούσαν να χρησιμοποιήσουν για την αποστολή των μηνυμάτων τους. Πολλά μηνύματα αυτής της κατηγορίας μεταφέρονται επισυναπτόμενα, τα οποία μπορεί να είναι **ιοί** ή **δούρειοι ίπποι**, οι οποίοι θέτουν σε κίνδυνο την ασφάλεια των συστημάτων. Το τελευταίο διάστημα μεγάλο ποσοστό ανεπιθύμητης και επικίνδυνης αλληλογραφίας είναι αποτέλεσμα της δράσης ιών που έχουν προσβάλει διάφορα συστήματα διασυνδεμένα στο Διαδίκτυο.

2.2.**Προστασία από ανεπιθύμητη αλληλογραφία(spam)**

2.2.1 Για τους απλούς χρήστες

- **Δεν πρέπει να δημοσιεύονται οι διευθύνσεις ηλεκτρονικού ταχυδρομείου.** Βάζοντας τη διεύθυνση ηλεκτρονικού ταχυδρομείου σε μια ιστοσελίδα είναι σχεδόν σίγουρο ότι σύντομα θα σταλούν στο χρήστη μηνύματα Spam στο γραμματοκιβώτιο.

- **Δε πρέπει να δίνεται η διεύθυνση ηλεκτρονικού ταχυδρομείου , σε οργανισμούς που δεν είναι εμπιστοσύνης.** Οι χρήστες πρέπει να είναι προσεκτικοί όταν επισκέπτονται διάφορους δικτυακούς τόπους και ζητείται η συμπλήρωση προσωπικών στοιχείων και στοιχείων επικοινωνίας, όπως το e-mail. Αν είναι αναγκαστικό να δοθεί η διεύθυνση ηλεκτρονικού ταχυδρομείου, πρέπει να διαβαστούν προσεκτικά οι όροι χρήσης και η πολιτική εχεμύθειας με την οποία δεσμεύεται ο συγκεκριμένος οργανισμός.

- **Δεν πρέπει να δίνεται απάντηση στο spam**. Οι χρήστες δε πρέπει να απαντάνε

στους spammers ακόμα και στην ένδειξη για διαγραφή από τις mail λίστες τους. Είναι μια παγίδα με τελικό αποτέλεσμα:

- Να διαπιστωθεί η εγκυρότητα της mail διεύθυνσης του χρήστη και επομένως να γίνει στόχος αποστολής επιπλέον μηνυμάτων.
- Να χαθεί χρόνος του χρήστη και να σπαταλήσει πόρους χωρίς λόγο, ενώ δεν υπάρχει αποτέλεσμα.

- **Αναφορά κάθε μηνύματος Spam που λαμβάνει ο χρήστης.** Υπάρχουν σχετικές υπηρεσίες του Διαδικτύου οι οποίες διατηρούν λίστες spammers. Τις λίστες αυτές αξιοποιούν πολλοί εξυπηρετητές ηλεκτρονικού ταχυδρομείου για τον περιορισμό του Spam που φθάνει στους χρήστες. Στις υπηρεσίες αυτές μπορεί ο χρήστης να αναφέρει τα μηνύματα τύπου Spam που φθάνουν σε αυτόν.
- **Διάδοση της γνώσης και της εμπειρίας σε σχέση με το Spam**. Ενημέρωση των χρηστών του δικτύου του κάθε χρήστη, μαθητές, εκπαιδευτικούς, διοικητικό προσωπικό, την οικογένεια του και τους φίλους του, για το θέμα του Spam και την αντιμετώπιση του. Είναι αρκετά συνηθισμένο οι spammers να συγκεντρώνουν e-mail διευθύνσεις από τις απαντήσεις χρηστών του Διαδικτύου.
- **Έλεγχος των συστημάτων του χρήστη, ώστε να είναι σωστά διαμορφωμένα και ασφαλή.** Ένα μεγάλο ποσοστό του Spam διαδίδεται από mail servers που δεν είναι σωστά διαμορφωμένοι (Open Relay), αλλά ακόμα και από συστήματα χρηστών.
- **Προγράμματα αλληλογραφίας με δυνατότητα εντοπισμού της ενοχλητικής αλληλογραφίας (Spam - Junk Email).** Τα σύγχρονα προγράμματα αλληλογραφίας όπως το Microsoft Outlook 2003, Mozilla Mail κλπ., επιτρέπουν αυτόματα τον εντοπισμό και το φιλτράρισμα του spam. Χρησιμοποιούν διάφορες τεχνικές, η πιο

γνωστή από τις οποίες είναι η Bayesian filtering και διάφορες παραλλαγές. Τα προγράμματα αυτά επιτρέπουν τον διαχωρισμό του spam σε ξεχωριστό φάκελο π.χ.: Junk Email στο Microsoft Outlook ή επιτρέπουν στον χρήστη την οριστική διαγραφή των μηνυμάτων που χαρακτηρίζονται ως Junk Email.

- **Χρήση white lists.** Τα ίδια προγράμματα επιτρέπουν στον χρήστη να συντηρεί στον υπολογιστή του black και white lists. Με την τεχνική αυτή επιτρέπεται η παραλαβή μηνυμάτων μόνο από τους χρήστες που υπάρχουν στην white list η οποία μπορεί να επεκτείνεται και στο βιβλίο διευθύνσεων που συντηρεί ο χρήστης στο πρόγραμμα ηλεκτρονικού ταχυδρομείου του.

- **Φιλτράρισμα με βάση τον αποστολέα και το περιεχόμενο.** Τα σύγχρονα προγράμματα ηλεκτρονικού ταχυδρομείου παρέχουν τη δυνατότητα στο χρήστη να ορίσει φίλτρα με βάσει το περιεχόμενο και τον αποστολέα και να αποφασίσει τι θα κάνει το μήνυμα αυτό. Έτσι ο χρήστης αποκλείει ανεπιθύμητους αποστολείς ενώ ταυτόχρονα μπορεί να κάνει φιλτράρισμα με βάση το περιεχόμενο του μηνύματος. Για παράδειγμα με δική του ευθύνη ορίζει κάθε μήνυμα με τη φράση *FREE LIVE PICTURE* να χαρακτηρίζεται ως spam και/ή να διαγράφεται ή να αποθηκεύεται σε ξεχωριστό φάκελο.

2.2.2. Μέθοδοι φιλτραρίσματος

Οι πιο συνηθισμένοι μέθοδοι φιλτραρίσματος με βάση το περιεχόμενο είναι :

- **Attachments**. Το μπλοκάρισμα κάποιων συγκειμένων τύπων αρχείων, για παράδειγμα το μπλοκάρισμα των εκτελέσιμων αρχείων.

- **Bayesian.** Είναι μια στατιστική τεχνική e-mail φιλτραρίσματος. Χρησιμοποιεί έναν ταξινομητή Byes για τον εντοπισμό ανεπιθύμητων μηνυμάτων ηλεκτρονικού

ταχυδρομείου .

- Φιλτράρισμα βασισμένο στον **DNS.**
- **Char-set**. Ο έλεγχος των χαρακτήρων (char) ,αν δηλαδή όλοι οι χαρακτήρες του μηνύματος είναι χαρακτήρες που υπάρχουν, είναι δηλαδή πραγματικοί, και συμφωνούν με τους χαρακτήρες του υπόλοιπου μηνύματος.
- **Content-Encoding**. Την κωδικοποίηση του μηνύματος.
- **Heurestic**. Φιλτράρισμα βασισμένο σε ευρετικό σκοράρισμα στο περιεχόμενο ,που στηρίζεται σε πολλαπλά κριτήρια.
- **HTML**. Έλεγχος για πιθανές ανωμαλίες στον κώδικα HTML.
- **Γλώσσα.** Φιλτράρισμα βάσει τη γλώσσα γραφής.
- **Mail Header.** Φιλτράρισμα που βασίζεται αποκλειστικά στην ανάλυση των e-mail κεφαλίδων.
- **Mailing list.** Χρησιμοποιείται για την ανίχνευση μηνυμάτων που περιέχουν λίστες e-mail.
- **Φράσεις.** Φιλτράρισμα με βάση την ανίχνευση φράσεων που περιέχονται στο κείμενο.
- **URL.** Φιλτράρισμα με βάση το URL. Κατάλληλο για κλείδωμα ιστοσελίδων ή τμήματα ιστοσελίδων.
- **Εξελιγμένα προγράμματα φιλτραρίσματος.** Υπάρχει μια σειρά από εξελιγμένα ελεύθερα και εμπορικά προγράμματα spam φίλτρων.
- **DNSBL (DNS-based Blackhole List**, **Block List**, **ή Blacklist).** Μια DNSBL λίστα είναι μια λίστα διευθύνσεων IP που δημοσιεύεται μέσω του Internet Domain Name Service, είτε σαν αρχείο που μπορεί να χρησιμοποιηθεί από κάποιο λογισμικό ενός

διακομιστή DNS, είτε ως μια ζωντανή ζώνη DNS που μπορεί ο χρήστης να ενημερωθεί σε πραγματικό χρόνο. Αυτές οι λίστες χρησιμοποιούνται για να δημοσιεύονται οι διευθύνσεις των υπολογιστών ή δικτύων που συνδέονται με το spamming. Τα πιο πολλά λογισμικά anti-spam για mail servers μπορούν να ρυθμιστούν ώστε να απορρίπτουν ή να σημαδεύουν e-mails που έχουν σταλεί από κάποιο site ή server το οποίο βρίσκεται σε μια ή περισσότερες τέτοιες λίστες. Υπάρχουν δεκάδες DNSBLs λίστες, τα οποία χρησιμοποιούν ένα ευρύ φάσμα κριτηρίων για την εγγραφή και τη διαγραφή των διευθύνσεων. Αυτές μπορεί να περιλαμβάνουν καταλόγους από διευθύνσεις από μολυσμένους υπολογιστές ή από άλλου είδους μηχανήματα που χρησιμοποιούνται για την αποστολή spam. Ωστόσο, αυτός ο τρόπος αντιμετώπισης δεν είναι αλάνθαστος. Οι διάφορες λίστες, αν και χρησιμοποιούν διαφορετικές μεθόδους εφαρμογής κάθε φορά, πάντα εγκυμονούν τον κίνδυνο κάποιο χρήσιμο μήνυμα από έγκυρο αποστολέα να γίνει αντιληπτό ως spam και συνεπώς να διαγραφεί (ή να ταξινομηθεί στον φάκελο spam που είναι εύκολο να μην του δώσουμε προσοχή). Τα περισσότερα συστήματα φιλτραρίσματος με βάση το περιεχόμενο χρησιμοποιούν ένα συνδυασμό τεχνικών.

2.2.3Περιορισμοί φίλτρων

Τα anti spam που βασίζονται σε φίλτρα έχουν τρείς σημαντικούς περιορισμούς:

Παρακάμπτοντας τα φίλτρα. Οι εφαρμογές των spammers και των *bulk mail* δεν είναι στατικές , δηλαδή μπορούν και προσαρμόζονται πολύ γρήγορα στα φίλτρα. Για παράδειγμα, για να αντιμετωπίσουν τις λίστες λέξεων, οι αποστολείς spam αλλάζουν με τυχαίο τρόπο

την ορθογραφία των λέξεων ,για παράδειγμα («*viagra*", "*V1agra*", "\ / *iaagra*"). Hash busters (ακολουθίες τυχαίων χαρακτήρων που διαφέρουν σε κάθε e-mail) δημιουργήθηκαν για την παράκαμψη φίλτρα hash. Σήμερα είναι πολύ δημοφιλή τα φίλτρα Bayesian. Τα περισσότερα φίλτρα spam είναι αποτελεσματικά μόνο για λίγες εβδομάδες στην καλύτερη περίπτωση. Προκειμένου να διατηρηθεί η βιωσιμότητα ενός φίλτρου anti-spam πρέπει να ενημερώνεται διαρκώς, συνήθως σε καθημερινή ή εβδομαδιαία βάση.

Ψευδώς θετικά. Όσο πιο αποτελεσματικό είναι ένα φίλτρο spam, τόσο μεγαλύτερη είναι η πιθανότητα λάθους εκτίμησης ενός επιθυμητού e-mail ως spam. Για παράδειγμα, το e-mail που περιέχει τη λέξη "*viagra*" (π.χ., το κείμενο που είναι spam "*Ελεύθερο* viagra*"* ή το προσωπικό email που δεν είναι spam "*Γεια σου, είδες την αστεία διαφήμιση του viagra?*") είναι σχεδόν βέβαιο ότι θα σημειωθεί ως spam ανεξάρτητα από το περιεχόμενο.

Φίλτρο αναθεώρηση. Τα μηνύματα που χαρακτηρίζονται ως spam δεν διαγράφονται, συνήθως αμέσως. Αντ 'αυτού, τα μηνύματα αυτά τοποθετούνται σε «γραμματοκιβώτια spam» για μελλοντική αναθεώρηση. Δυστυχώς, αυτό σημαίνει ότι οι χρήστες εξακολουθούν να βλέπουν το spam, έστω και αν βλέπουν μόνο το θέμα του μηνύματος, καθώς ψάχνουν στους φακέλους spam για τυχόν σωστά e-mails. Στην ουσία το φιλτράρισμα βοηθάει στη διαλογή εισερχομένων μηνυμάτων.

2.2.4.Μειονεκτήματα

Το μειονέκτημα των πιο πάνω λύσεων και με δεδομένο τον τρόπο λήψης της αλληλογραφίας από τους χρήστες (POP3 και IMAP), είναι ότι η δράση αυτών των φίλτρων απαιτεί την λήψη της αλληλογραφίας στο γραμματοκιβώτιο του χρήστη και την εφαρμογή των κανόνων μετά από αυτό. Αυτό επιβαρύνει την γραμμή του χρήστη και προκαλεί

εκνευρισμό όταν μέσω απλής σύνδεσης PSTN ο χρήστης πρέπει να λάβει και τον όγκο της αλληλογραφίας Spam. Υπάρχουν φίλτρα τα οποία μπορούν να εφαρμοστούν στην πλευρά του mail server πριν την είσοδο των μηνυμάτων στο γραμματοκιβωτίου του χρήστη, αλλά για την εφαρμογή απαιτούν υποστήριξη από τον παροχέα υπηρεσιών Διαδικτύου, ο οποίος μπορεί να παρέχει την εν λόγω προστασία με όρους και προϋποθέσεις που αυτός επιλέγει.

2.2.5.Για τους διαχειριστές εξυπηρετητών ηλεκτρονικού ταχυδρομείου (mail servers)

Υπάρχουν διάφορες συσκευές, υπηρεσίες και συστήματα λογισμικού όπου οι διαχειριστές μπορούν να χρησιμοποιήσουν για τη μείωση του spam στα συστήματά τους. Κάποια από αυτά βασίζονται στο να απορρίπτουν e-mail από γνωστά ή πιθανά sites που στέλνουν spam. Άλλες πιο προηγμένες τεχνικές ανάλυσης που πραγματοποιούνται από τη μεριά του mail server, γίνονται σε πραγματικό χρόνο(real time) και τη συγκρίνει τα μηνύματα με διάφορες παγκόσμιες βάσεις δεδομένων που περιέχουν πληροφορίες για spam. Πολλά συστήματα φιλτραρίσματος χρησιμοποιούνε τεχνικές μηχανικής μάθησης, που βελτιώνουν την ακρίβεια στους χειρισμούς των spam e-mails.

Οι βασικές λύσεις που εφαρμόζουν οι διαχειριστές ηλεκτρονικού ταχυδρομείου συγκεντρώνονται στα ακόλουθα σημεία:

- **Έλεγχος εγκυρότητας στο DNS και στους headers**. Απορρίπτονται τα μηνύματα που προέρχονται από εξυπηρετητές e-mail που δεν έχουν έγκυρες δηλώσεις DNS. Αλλες φορές απορρίπτονται μηνύματα αυτόματα ,όταν το domain στην e-mail

διεύθυνση (το τμήμα μετά το @) δεν υπάρχει στο DNS.

- **Χρήση SMTP Server που απορρίπτει γνωστούς spammers**. Ο διακομιστής εισερχόμενης αλληλογραφίας (SMTP) αρνείται να λάβει και απορρίπτει τα μηνύματα που προέρχονται από Servers που διακινούν Spam ή δεν ικανοποιούν τις προδιαγραφές ασφαλείας και άρα μπορούν να χρησιμοποιηθούν από τους Spammers για την διακίνηση μηνυμάτων. Η λύση αυτή βασίζεται στην χρήση DNSBL λιστών που αναφέραμε προηγουμένως και η οποία είναι μια Διεθνής πρακτική που εφαρμόζεται από πολλούς παροχείς υπηρεσιών e-mail.
- **Χρήση προγραμμάτων προστασίας στον διακομιστή**. Μία σειρά από λύσεις προστασίας στον διακομιστή με φίλτρα και εξελιγμένες τεχνικές (πχ Bayesian filtering) μπορούν να εφαρμοστούν. Στην κατεύθυνση αυτή υπάρχουν σχετικά εργαλεία εμπορικά και της φιλοσοφίας Ελεύθερου Λογισμικού που είναι διαθέσιμα και μπορούν να χρησιμοποιηθούν.
- **Φιλτράρισμα των SMTP συνδέσεων.** Εφαρμογή φίλτρων που δεν επιτρέπουν σύνδεση στους διακομιστές αλληλογραφίας από γνωστούς εξυπηρετητές που διακινούν Spam.
- **Παρακολούθηση.** Η παρακολούθηση της κίνησης e-mail με στόχο τον εντοπισμό της αποστολής μεγάλου αριθμού μηνυμάτων από συγκεκριμένους αποστολείς οι οποίοι θα μπορούσαν να είναι spammers ειδικά αν το φαινόμενο είναι επαναλαμβανόμενο.

3.Αλγόριθμοι HASH

3.1.Αλγόριθμοι hash

Μια **συνάρτηση κατακερματισμού (hash function)** είναι μια σαφώς καθορισμένη διαδικασία ή μαθηματική συνάρτηση που μετατρέπει μια μεγάλη, ίσως Variable-sized, ποσότητα δεδομένων σε ένα μικρό datum. Οι τιμές που επιστρέφονται από μια συνάρτηση κατακερματισμού ονομάζονται **τιμές hash, κωδικοί hash ,hash sums, check sums** ή απλά **hashes.** Η ιδανική κρυπτογραφική συνάρτηση κατακερματισμού έχει τέσσερις βασικές ιδιότητες:

- Να είναι εύκολος ο υπολογισμός της τιμής hash οποιουδήποτε μηνύματος.
- Είναι ανέφικτο ένα μήνυμα να έχει μια δεδομένη τιμή hash.
- Είναι ανέφικτο να τροποποιηθεί ένα μήνυμα χωρίς να αλλάξει η τιμή.
- Είναι ανέφικτο να βρεθούν δύο διαφορετικά μηνύματα με το ίδιο κωδικό hash.

3.2.Αλγόριθμος MD5 hash

Στην κρυπτογραφία, ο αλγόριθμος **MD5 (Message-Digest αλγόριθμος 5)** είναι μια ευρέως χρησιμοποιούμενη κρυπτογραφική συνάρτηση hash, με 128 - bit (16-byte) τιμή κατακερματισμού. Ο MD5 έχει χρησιμοποιηθεί σε ένα ευρύ φάσμα εφαρμογών ασφάλειας, και χρησιμοποιείται επίσης ευρέως για να ελέγχει την ακεραιότητα των αρχείων . Ωστόσο, έχει αποδειχθεί ότι ο MD5 δεν είναι ανθεκτικός στις συγκρούσεις, σαν αποτέλεσμα ο MD5 δεν είναι κατάλληλος για εφαρμογές όπως SSL πιστοποιητικά ή ψηφιακές υπογραφές. Ο MD5 hash συνήθως εκφράζεται ως ένα 32-ψήφιο δεκαεξαδικό αριθμό.

4.Ανάλυση Συστήματος Anti-SPAM

4.1.Γενικές πληροφορίες συστήματος

Όπως αναφέρθηκε τα πιο πολλά εργαλεία αντιμετώπισης spam δουλεύουν στο γραμματοκιβώτιο του χρήστη και ως αποτέλεσμα έχει την επιβάρυνση της γραμμής του χρήστη όσο και τον πολύτιμο χρόνο του ιδίου. Πέρα από όλα τα εργαλεία που κυκλοφορούν στην αγορά ο καλύτερος τρόπος για να κριθεί εάν ένα e-mail είναι spam ή όχι, είναι ο ίδιος χρήστης. Σκοπός του Anti-Spam συστήματος είναι, με τη βοήθεια των χρηστών, η μεταφορά αντιμετώπισης του spam στους διαχειριστές εξυπηρετητών ηλεκτρονικού ταχυδρομείου (mail servers).

Χρησιμοποιώντας τη γλώσσα προγραμματισμού J**ava** υλοποιήθηκε ένας Mail Client, ο οποίος μπορεί να στέλνει και να δέχεται e-mails, να αποθηκεύει τα αρχεία (attachments) των εισερχομένων e-mails. Ο Mail Client εκτός από τις συνηθισμένες επιλογές (New Message, Reply , Forward, Delete) έχει επιπλέον λειτουργίες και κάποια επιπλέον πεδία.

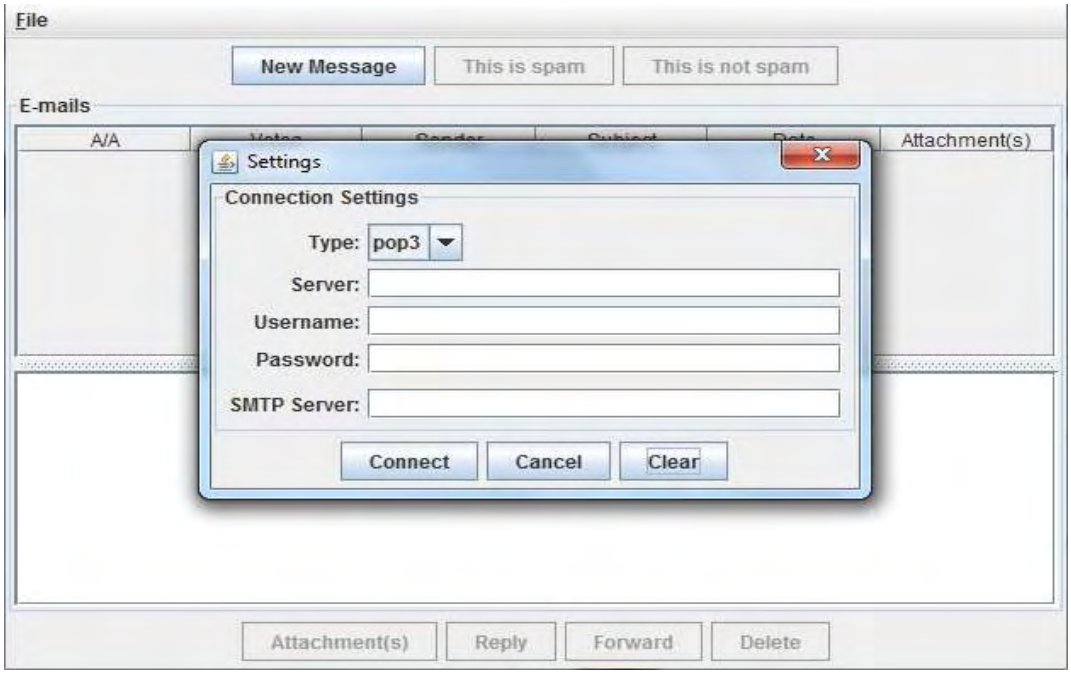
Στην εφαρμογή υπάρχει επιλογή "*This is spam*" με την οποία ο χρήστης μπορεί να ψηφίσει αν κάποιο e-mail το θεωρεί spam. Η ψήφος του, χρησιμοποιώντας τον Hash αλγόριθμο MD5, αποθηκεύεται σε μια βάση δεδομένων(ΒΔ) σαν κωδικός hash του e-mail που ψήφισε. Υπάρχει επίσης πεδίο (Votes) όπου παρουσιάζει πόσοι χρήστες έχουν ψηφίσει για αυτό το e-mail. Σε περίπτωση που ο χρήστης αλλάξει γνώμη για το αν το e-mail είναι spam υπάρχει η επιλογή "*This is not spam*" όπου και αναιρείται η ψήφος του.

Με αυτό το τρόπο οι διαχειριστές των mail servers, με τη βοήθεια της ΒΔ θα μπορούν να "κόβουν" τα e-mails που έχουν συγκεντρώσει ικανοποιητικό αριθμό ψήφων

και να μην τα προωθούν ξανά. Έτσι η αντιμετώπιση των spam mails θα φύγει κατά ένα μεγάλο μέρος από τον χρήστη.

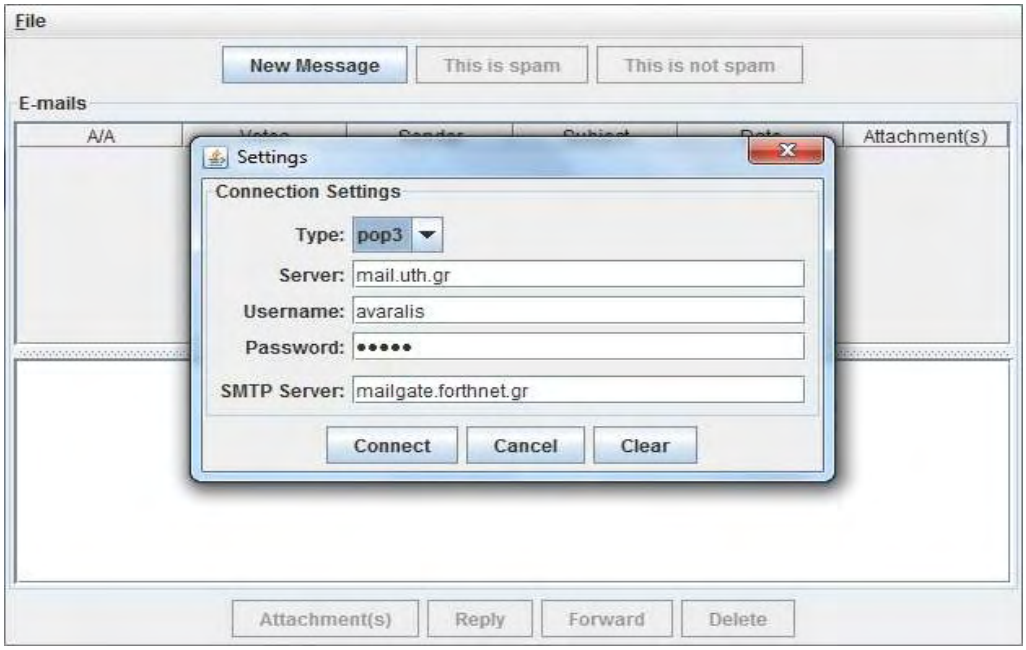
4.2.Αναλυτική περιγραφή εφαρμογής

Πιο αναλυτικά, ανοίγοντας το πρόγραμμα εμφανίζεται το παράθυρο Settings(Εικόνα 1). Ο χρήστης στο πρώτο πεδίο μπορεί να επιλέξει πρωτόκολλο παραλαβής αλληλογραφίας POP3 ή IMAP. Στα επόμενα πεδία (Server) συμπληρώνει τον server εισερχόμενης αλληλογραφίας, το όνομα του χρήστη(Username), τον κωδικό (Password) και τέλος τον server εξερχόμενης αλληλογραφίας (SMTP Server).Το κουμπί Clear 'καθαρίζει' τα στοιχεία από προηγούμενο χρήστη.



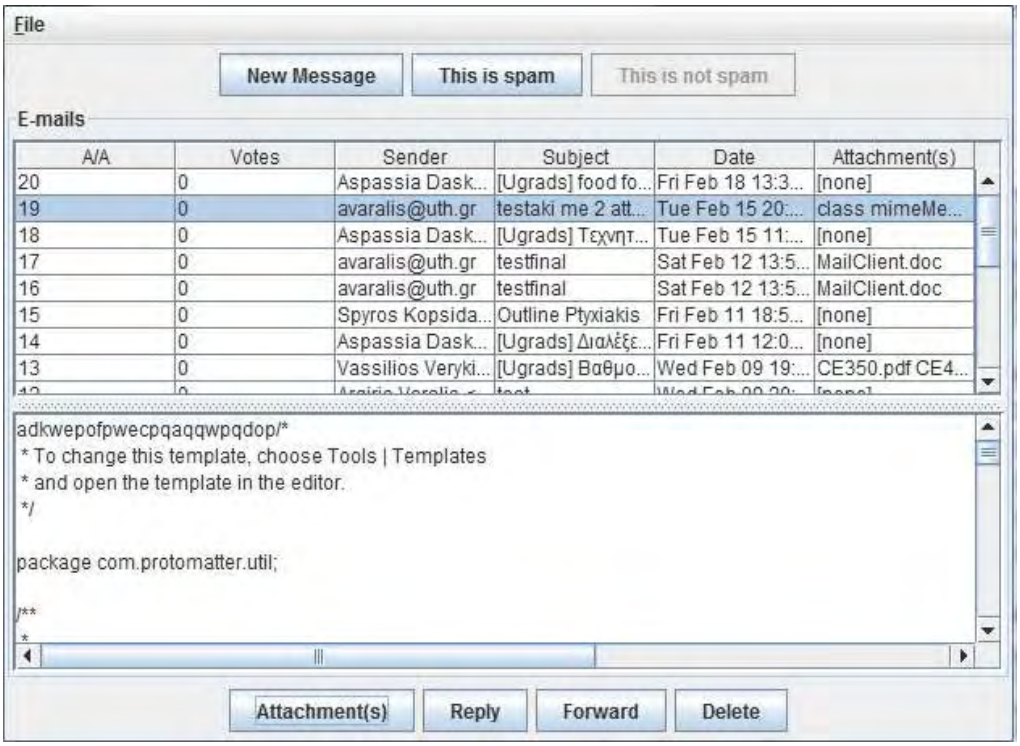
Εικόνα 1 : Έναρξη προγράμματος

Ένα παράδειγμα συμπλήρωσης.



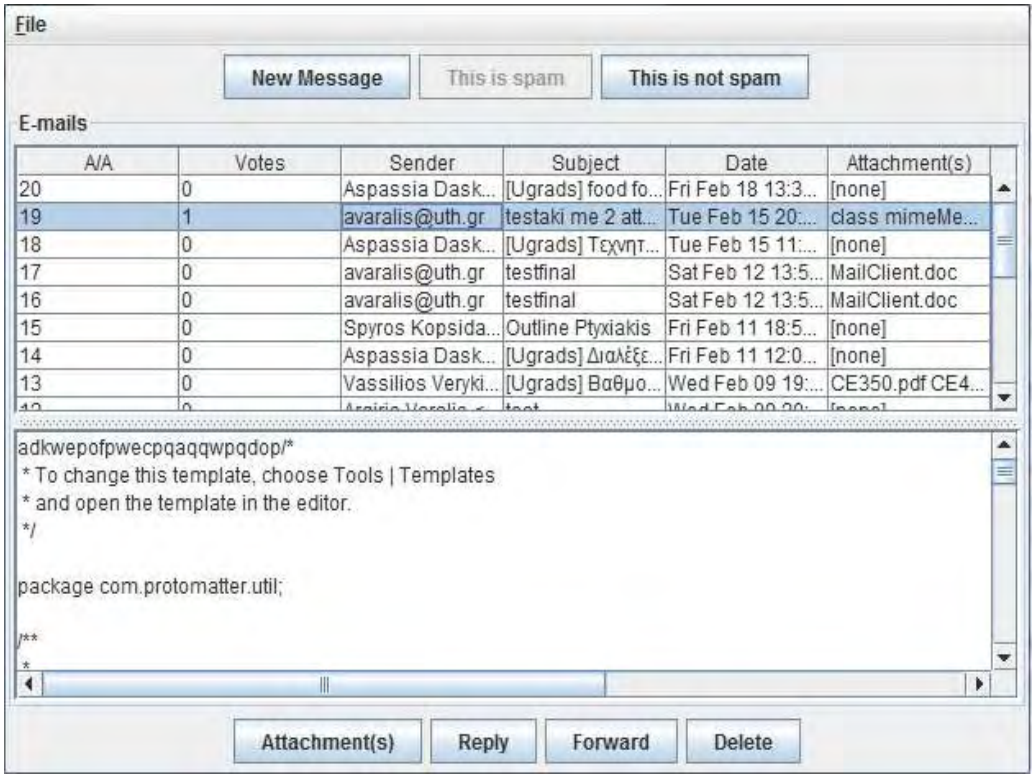
Εικόνα 2 : Αρχικό παράθυρο Settings ,συμπληρωμένο

Πατώντας το κουμπί Connect στην Εικόνα 2 γίνεται η σύνδεση και εμφανίζονται τα εισερχόμενα e-mail(Εικόνα 3)l.Κάνοντας κλίκ σε κάποιο e-mail το περιεχόμενο του εμφανίζεται στο κενό από κάτω από τα εισερχόμενα e-mail. Αν το e-mail έχει κάποιο attachment τότε το κουμπί Attachment(s) είναι ενεργό και επιλέγοντάς το, ο χρήστης μπορεί να αποθηκεύσει αυτά τα αρχεία.



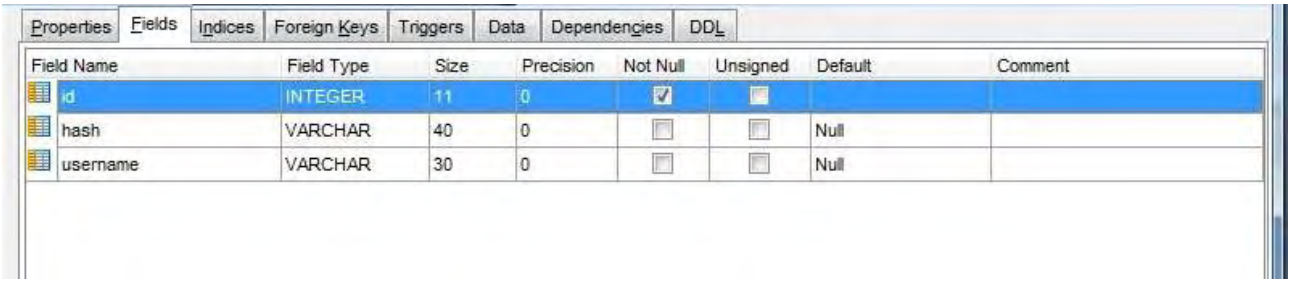
Εικόνα 3 :Στιγμιότυπο λειτουργίας του προγράμματος

Αν ο χρήστης θεωρεί ότι κάποιο e-mail είναι spam απλά επιλέγει το κουμπί "*This is spam*". Στην Εικόνα 4 έχει επιλεχτεί ένα μήνυμα ως spam.Το πεδίο *Votes* του e-mail που επιλέχθηκε αυξάνεται κατά ένα και ο MD5 κωδικός hash αποθηκεύεται στη ΒΔ .Το κουμπί "*This is spam*" απενεργοποιείται και ενεργοποιείται το κουμπί "*This is not spam*".

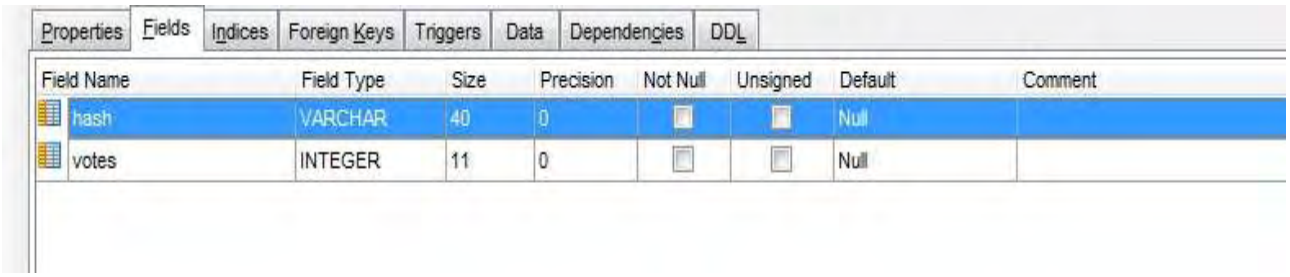


Εικόνα 4 : Στιγμιότυπο μετά την ψήφο ενός e-mail.

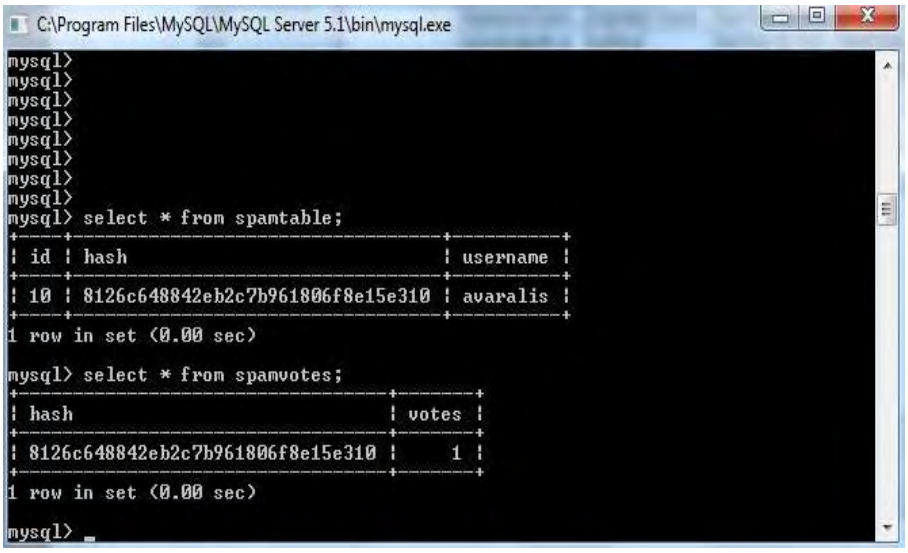
Στη ΒΔ(Εικόνα 5) υπάρχουν δύο tables. Στο *spamtable* αποθηκεύεται το id(αριθμός καταχώρησης) το Username , κωδικός hash και στο *spamvotes* αποθηκεύεται ο κωδικός hash και ο αριθμός των ψήφων του συγκεκριμένου κωδικού hash.



Εικόνα 5 :Το spamtable και τα περιεχόμενα στοιχεία του



Εικόνα 6 : Το spamvotes και τα περιεχόμενα στοιχεία του



Εικόνα 5 : Βάση Δεδομένων(ΒΔ),μετά την ψήφο

Αν ο χρήστης αλλάξει γνώμη για το αν το e-mail είναι spam μπορεί να πάρει πίσω τη ψήφο του επιλέγοντας το κουμπί "*This is not spam*". Έτσι το πεδίο *Votes* μειώνεται κατά ένα κατά ένα και διαγράφονται οι εγγραφές από το table *spamtable* και μειώνεται κατά ένα η εγγραφή από το *spamvotes* εκτός αν είναι μηδέν που σβήνεται τελείως η εγγραφή.

5.Μελλοντικές Επεκτάσεις

Η ιδέα των ψήφων στη ΒΔ θα μπορούσε μελλοντικά να επεκταθεί και να χρησιμοποιείται από τους διαχειριστές των mail servers. Ένα παράδειγμα θα ήταν η δημιουργία μιας κεντρικής ΒΔ στο Πανεπιστήμιο Θεσσαλίας όπου ο mail server θα ενημερώνεται από αυτή και όλοι οι χρήστες θα την ενημερώνουν. Σαν αποτέλεσμα τα spam e-mails θα σταματούσαν να προωθούνται στους χρήστες σε πολύ γρήγορο χρονικό διάστημα.

Άλλη μελλοντική επέκταση θα ήταν η δημιουργία ενός καλύτερου GUI της εφαρμογής, με πιο πολλές επιλογές στο μενού, αλλά κρατώντας την απλή και εύχρηστη λειτουργία προς το χρήστη.

6. Συμπεράσματα

Υπάρχουν διάφορες απόψεις στη δικτυακή κοινότητα. Κάποιοι πιστεύουν ότι το spam δε μπορεί να σταματήσει ποτέ, κάποιοι ότι το spam είναι ευθύνη των τελικών χρηστών και κάποιοι άλλοι ότι είναι ευθύνη των διαχειριστών mail servers. Η αλήθεια είναι ότι το spam αυξάνεται συνεχώς. Παρόλο που δημιουργούνται νέα και ανανεώνονται τα υπάρχοντα μέσα για την αντιμετώπιση του spam, οι spammers πάντα βρίσκουν νέους τρόπους για

καταφέρουν τον σκοπό τους. Οι χρήστες βλέπουν μόνο ένα μικρό ποσοστό spam στο mailbox τους, δεδομένου ότι οι κατάλογοι των spammers περιέχουν συχνά ένα μεγάλο ποσοστό από άκυρα emails και πολλά φίλτρα spam διαγράφουν απλά ή απορρίπτουν το «προφανές spam». Μια έρευνα του 2010 για τους χρήστες του ηλεκτρονικού ταχυδρομείου σε Ευρώπη και ΗΠΑ , έδειξε ότι παρόλο που γνωρίζανε τους κινδύνους των spam mails, το 46% εξακολουθεί να τα ανοίγει και να βάζουν τους υπολογιστές τους σε κίνδυνο. Άρα από όλα αυτά βγαίνει το συμπέρασμα πως για να περιοριστεί το spam πρέπει όλοι οι χρήστες αλλά και οι διαχειριστές των mail servers από κοινού να συμβάλλουν σε αυτό.

7.Βιβλιογραφία

Για τον κώδικα της πτυχιακής χρησιμοποιήθηκε σαν βάση το ακόλουθο site :

Java Tips-How to create an e-mail client in Java	http://www.java-tips.org/java-se-tips/javax.swing/how-to-create-an-e-mail-client-in-java.html

Για τη βιβλιογραφία της πτυχιακής χρησιμοποιήθηκαν τα ακόλουθα sites :

Κέντρο Πληροφορικής & Νέων Τεχνολογιών Ηλείας	http://dide.ilei.sch.gr/keplinet/tech/spam.php
Πανελλήνιο Σχολικό Δίκτυο	http://www.sch.gr/2010-04-07-09-22-34/-spam/%CE%A3%CE%B5%CE%BB%CE%AF%CE%B4%CE%B1-2

Πανελλήνιο Σχολικό Δίκτυο	http://www.sch.gr/sch-portlets/static/manual/aboutSpam/index.php?_list=opinions
Wikipedia	http://en.wikipedia.org/wiki/DNSBL
Wikipedia	http://en.wikipedia.org/wiki/Anti-spam_techniques
Symatec	http://www.symantec.com/connect/articles/anti-spam-solutions-and-security
Reason.com	http://reason.com/archives/2003/11/01/the-spam-wars